LOCKHEED MARTIN

GCCS-A Data Synchronization

29 April 2002

LOCKHEED MARTIN

Background

- Some Confusion Exists on The Method Used for Synchronizing Data
- Objective is Effectively Explain How and Why

LOCKHEED MARTIN

Background

- The GCCS-A System consists of multiple servers tied together using Informix Enterprise Replication and the RepSync Application. The system should be thought of as the collection of all servers that are tied together versus a single server being a system
- A common and consistent rule for resolving data differences is required across the entire system
- Both Informix Enterprise Replication and RepSync need to use a consistent rule for resolving data differences or they will work against each other

LOCKHEED MARTIN

Background

- The synchronization process for GCCS-A was designed based on the following:
  - Data synchronization must be automated. Sufficient resources to manually evaluate and resolve each row in each table for each server do not exist
  - Current data always takes precedence over older data. This rule supports automation
  - The most current data can originate on any server in the system
  - The most current data for various tables and rows can be mixed across all servers. That is, each server may contain some of the most current data

LOCKHEED MARTIN

Background

- The synchronization process for GCCS-A was designed based on the following (continued):
  - A master server with the correct and current data for every row and table does not exist
  - The system must be available for usage during synchronization
  - IER must continue to function during synchronization
  - Data that is fixed should not be further replicated to prevent network “flooding”
  - For RepSync, it is only practical to compare 2 servers (databases) at a time to determine if data is different

LOCKHEED MARTIN

Focus Server

- The term FOCUS server does not mean MASTER or AUTHORITATIVE server
- FOCUS server simply means that the majority of the code is executed there. The PES is a good choice because it usually has more processing capacity and better network capacity. If data is more recent (including more recent deletes) then that data wins
- The FOCUS server selected has no impact on the data comparison.

Examples (slightly tedious)

- **Data is compared for the agdb_data_v1 database, br_dm table for the PES and LAN servers.**
- **If the same row is found on both servers, with the same timestamp then no action is needed**
- **If the same row is found on both servers and the timestamp is newer on the LAN, then update the row on the PES**
- **If a row is found on the PES server and not the LAN server then check the delete log on the LAN server for a record of the delete**
  - If the data was deleted on the LAN server more recently than the timestamp on the PES server then delete the row from the PES
  - If there is no record of the delete on the LAN server, then put the row on the LAN server to match the PES

LOCKHEED MARTIN

Consider

- If we change RepSync to synchronize data based on a master server (where all the data has been determined to be correct for all tables for each row) what will happen?
- Newer data on the Non-Master server could be lost, unless data changes are not allowed on any server except the Master server
- IER would need to be modified to only replicate in a single direction from the designated Master server or to always let the Master server data to win in a conflict situation
- If we allow RepSync to have a different rule than IER, then IER over time will change the data back to before RepSync executed